VERALLGEMEINERUNG EINES VERFAHRENS VON L.KLEIN ZUR SCHÄTZUNG VON "DISTRIBUTED-LAG" MODELLEN

von

Peter FLEISSNER

Forschungsbericht Nr.36

Mai 1969

1. Einleitung

In den letzten Jahren kommen in ökonometrischen Modellen häufig nicht nur solche Gleichungen vor, die **Variable** enthalten, die keine, eine oder zwei Perioden verzögert auftreten, sondern auch Gleichungen, in denen eine abhängige Variable durch die Summe der gewichteten Werte der unabhängigen Variablen X_t, $_{t-1}$, ...X_{t-n}, ... aller vergangenen Perioden erklärt wird. Gleichungen dieses Typs nennt man "distributed lag"-Modelle. (Lit. 1)
In einer Formel ausgedrückt heißt das:

$$Y_t = \alpha + \beta \sum_{\tau=0}^{\infty} \lambda_\tau X_{t-\tau} + \varepsilon_t \qquad (1)$$

α, β sind reelle Zahlen, ε_t bedeutet eine Störgröße, die auf zwei verschiedene Arten interpretiert werden kann:

a) Die Gleichung (1) ist eine Strukturgleichung, d.h. X_t und Y_t sind exakt beobachtbare Größen, die exakte lineare Beziehung zwischen X und Y unterliegt aber einer Störung, die durch das additive Glied ε ausgedrückt wird.

b) Die Gleichung (1) stellt eine funktionale Beziehung dar, die immer exakt gilt. Allerdings ist Y_t nicht der wahre Wert der Variablen Y, sondern $E_t = Y_t - \varepsilon_t$. Daher kann man (1) auch schreiben:

$$E_t = \alpha + \beta \sum_{\tau=0}^{\infty} \lambda_\tau X_{t-\tau} \qquad (2)$$

Bei der Messung von Y begeht man also einen Fehler $\varepsilon_t = Y_t - E_t$.

2. Eine Spezialisierung des Modelles, Koyck-Transformation

Wählt man in (1) oder (2) die λ_τ so, daß $\lambda_\tau = \lambda^\tau$ wird, stellen die λ_τ also eine geometrische Folge dar, kann dieser Sonderfall besonders einfach behandelt werden.

$$Y_t = \alpha + \beta \sum_{\tau=0}^{\infty} \lambda^\tau X_{t-\tau} + \varepsilon_t \tag{3}$$

Da die Reihe $\sum_{\tau=0}^{\infty} \lambda^\tau X_{t-\tau}$ nur dann ökonomisch sinnvolle Bedeutung hat, wenn sie konvergiert, muß man neben der Beschränktheit der Werte von X verlangen, daß $|\lambda| < 1$.

Zur einfacheren Schreibweise wird nun ein Operator L, der sogenannte "lag"-Operator, eingeführt. Der Operator besitzt die Eigenschaft der Linearität

$$L(X_t^1 + X_t^2) = LX_t^1 + LX_t^2$$

und stellt eine Vorschrift dar, die eine Variable, auf die er angewandt wird, um eine Periode in die Vergangenheit verschiebt.

$$LX_t = X_{t-1}, \qquad L^k X_t = X_{t-k}.$$

Gleichung (3) erhält nun folgendes Aussehen:

$$Y_t = \alpha + \beta \sum_{\tau=0}^{\infty} \lambda^\tau L^\tau X_t + \varepsilon_t = \alpha + \beta \sum (\lambda L)^\tau X_t + \varepsilon_t =$$

$$= \alpha + \beta \frac{1}{1 - \lambda L} X_t + \varepsilon_t$$

Multiplikation mit $(1 - \lambda L)$ ergibt:

$$(1 - \lambda L) Y_t = (1 - \lambda L)\alpha + \beta X_t + (1 - \lambda L)\varepsilon_t \tag{4}$$

oder

$$Y_t = \lambda Y_{t-1} + \beta X_t + (1 - \lambda)\alpha + \varepsilon_t - \lambda \varepsilon_{t-1} \tag{5}$$

Der Übergang von Gleichung (3) auf (4) heißt Koyck-Transformation (Lit. 2). Die Koeffizienten in (3) sind nicht schätzbar, da in der Praxis keine beliebig langen Zeitreihen für das Wirtschaftsgeschehen existieren. Durch die Koyck-Transformation wurde Gleichung (3) in eine Form gebracht, die sogar mit Ordinary Least Squares (OLS) behandelt werden kann, wenn die Störgröße ε_t einem einfachen autoregressiven Schema

$$\varepsilon_t = \lambda\,\varepsilon_{t-1} + \eta_t$$

$$E(\eta_t) = 0 \qquad E(\eta_t\,\eta_{t-k}) = \begin{cases} \sigma_\eta^2 & \text{für } k = 0 \\ 0 & \text{für } k \neq 0 \end{cases}$$

gehorcht. Für diesen allerdings sehr speziellen Fall liefert OLS erwartungstreue Schätzwerte für die Parameter.

3. Ökonomische Interpretation

Die bisherigen, eher formalen Überlegungen bedürfen natürlich einer Entsprechung in der ökonomischen Wirklichkeit.
Das Problem stellt sich zunächst so: welche Verhaltensformen führen überhaupt zu Gleichungen der Gestalt (4) ?

Ich gebe hier zwei mögliche Interpretationen wieder, die beide mit dem Begriff des erwarteten Wertes einer ökonomischen Variablen operieren.

a) "Adaptive Expectations"-Modell (Lit. 3)

Einstweilen werden nur Annahmen über die erwarteten und tatsächlichen Werte der unabhängigen Variablen getroffen:
die unabhängige Variable X unterliege folgender Gleichung

$$X_t^+ - X_{t-1}^+ = \rho\,(X_{t-1} - X_{t-1}^+) \qquad (6)$$

wobei X_t^+ den erwarteten Wert zur Zeit t bedeutet, X_t den tatsächlichen Wert. ρ ist der Koeffizient der Erwartungen.

Er drückt den permanenten Anteil des Irrtums in der Erwartung zur Zeit (t-1) aus, der die Differenz der erwarteten Werte zweier aufeinanderfolgender Zeitpunkte erklärt. War z.B. der erwartete Umsatz einer Firma im Vorjahr kleiner als der tatsächliche, so wird heuer ein größerer Umsatz erwartet als im Vorjahr. ϱ ist ein Maß dafür, einen um wieviel größeren Umsatz man erwartet.

Im zweiten Schritt wird eine Beziehung zwischen der abhängigen Variablen Y und den erwarteten Werten von X postuliert.

$$Y_t = \alpha + \beta \; X_t^+ \qquad (7)$$

Wird X_t^+ aus (6) errechnet und in (7) eingesetzt, ferner X_{t-1}^+ aus (7) errechnet, so erhält man mit einem Fehlerterm v_t

$$Y_t = Y_{t-1} \; (1 - \varrho) + X_{t-1} \, \varrho \, \beta + \varrho \, \alpha + v_t \qquad (8)$$

Gleichung (8) stimmt bis auf die andere Nomenklatur mit (5) überein. Über die Fehlerelemente in den Gleichungen wird weiter unten diskutiert.

b) <u>"Partial Adjustment"</u> Modell

Dieses Modell führt ebenfalls zur Gleichung (8).
Hier werden Annahmen über die abhängige Variable Y getroffen.

$$Y_t - Y_{t-1} = \gamma \, (Y_{t-1}^+ - Y_{t-1}) \qquad 0 < \gamma < 1; \qquad (9)$$

d.h. die Differenz der aufeinanderfolgenden tatsächlichen Werte der abhängigen Variablen in 2 Perioden wird durch einen Anteil des Fehlers in den Erwartungen erklärt. γ, der Anpassungskoeffizient, stellt ein Maß für die Anpassung, z.B. der Produktionsmengen bei vorhandenen technischen Möglichkeiten und institutionellen Gegebenheiten dar.

Außerdem sei der erwartete Wert von Y durch irgendeine andere Variable X bestimmt:

$$Y_t^+ = bX_t + \alpha \tag{10}$$

(9`, (10) und ein Fehlerterm v_t ergeben

$$Y_t = (1 - \mu)\, Y_{t-1} + \mu b X_{t-1} + \alpha\mu + v_t \tag{11}$$

Die beiden Modelle (8) und (11) können statistisch nicht voneinander unterschieden werden. (Lit. 3)

4. Fehlermodelle

Identität der Gleichungen (5), (8) und (11) erreicht man erst, wenn die Fehlermodelle angegeben werden, denen die Gleichungen (6), (7) bzw. (9), (10) unterliegen. Wie leicht nachgeprüft werden kann, erhält man für v_t den Ausdruck $\varepsilon_t - \varepsilon_{t-1}$, wenn (6) exakt gilt, zu rechten Seite von (7) aber der Fehlerterm ε_t addiert wird.

Im "Partial Adjustment" Modell bleibt (1C) unverändert, aber Gleichung (9) gilt exakt für die wahren Werte von Y, also für E_t (siehe auch Abschnitt 1.b)

$$E_t - E_{t-1} = \mu\, (Y_{t-1}^+ - E_{t-1}) \tag{12}$$

Die wahren Werte von Y sind der Messung nicht zugänglich, kommen also auch in der Gleichung (11),die geschätzt werden soll, nicht vor.

5. Verallgemeinerung des KLEIN'schen Schätzverfahrens für Distributed Lag Modelle

Für den Fall einer einzigen unabhängigen Variablen hat KLEIN (Lit. 4) ein einfaches und elegantes Schätzverfahren angegeben, das für den Fall von nicht autokorrelierten Residuen der Gleichung (3) konsistent, für den Fall autokorrelierter Residuen zwar nicht konsistent, aber noch brauchbar ist (Lit. 5).

Es handelt sich bei diesem Schätzverfahren um einen Maximum likelyhoodestimator, der identisch mit dem Problem ist,

$$\sum_{t=1}^{N} \varepsilon_t^2 + \sum_{t=1}^{N} \varepsilon_{t-1}^2 \quad \text{bezüglich } \alpha, \beta, \lambda \quad \text{und } E_t \text{ zu minimieren.}$$

Die Varianz der Residuen wird bei diesem Verfahren nicht konsistent geschätzt, die Schätzgröße läßt sich aber leicht durch einen Faktor in eine konsistente verwandeln (Lit. 6).

Die asymptotischen Eigenschaften der Varianz-Kovarianz-Matrix der Koeffizienten werden in Lit. 7 angegeben.

Nimmt man nun als Arbeitshypothese an, die abhängige Variable Y wäre nur mit einem Fehler meßbar, der einer Gauß-Verteilung gehorcht, die wahren Werte E_t sind daher unbekannt, und die abhängige Variable Y^+ werde durch eine Linearkombination von mehreren Variablen erklärt (siehe Gleichung (12) und (10), Verallgemeinerung des Partial Adjustment Modells), so würde die Schätzung einer solchen Gleichung mit OLS verzerrte Ergebnisse liefern (Lit. 8), da die Residuen nicht von allen Regressoren unabhängig sind.

Erweitert man aber das oben angeführte Klein'sche Verfahren, so ist eine konsistente Schätzung der Parameter möglich. Die verallgemeinerte Gleichung lautet nun:

$$Y_t = \lambda Y_{t-1} + \sum_{i=1}^{M} \beta_i X_{i,t-1} + \alpha + v_t$$

$$v_t = \varepsilon_t - \lambda \varepsilon_{t-1} \qquad E_t = Y_t - \varepsilon_t \qquad (13)$$

Dieselbe Gleichung erhält man, wenn im Partial Adjustment Modell anstelle von (6) ein Gleichungssystem von verschiedenen unabhängigen Variablen mit dem gleichen ρ tritt und Gleichung (7) auf eine Linearkombination der erwarteten Werte der unabhängigen Variablen erweitert wird. Als Fehlermodell kann das gleiche wie beim Adaptive Expectations Modell dienen.

$$X_{i,t}^{+} - X_{i,t-1}^{+} = \rho\,(X_{i,t-1} - X_{i,t-1}^{+}) \qquad i = 1,2\ldots M \qquad (6')$$

$$Y_t = \alpha + \sum \beta_i X_{i,t}^{+} + \varepsilon_t \qquad (7')$$

6. Berechnung der Schätzgrößen der Parameter im verallgemeinerten Modell

Zur leichteren Rechnung gehen wir von den Originaldaten zu Abweichungen vom jeweiligen Mittelwert über.

Aus Gleichung (13) wird mit

$$\overline{X}_i = \frac{1}{N-1} \sum_{t=2}^{N} X_{i,t} \qquad X_{i,t} = \overline{X}_i + x_{i,t} \qquad i = 1,2\ldots M$$

$$\overline{Y} = \frac{1}{N-1} \sum_{t=2}^{N} Y_t \qquad Y_t = \overline{Y} + y_t$$

$$\overline{Y}_{-1} = \frac{1}{N-1} \sum_{t=2}^{N} Y_{t-1} \qquad Y_{t-1} = \overline{Y}_{-1} + y_{-1,t-1}$$

$$y_t = \lambda\, y_{-1,t-1} + \sum_i \beta_i\, x_{i,t} + v_t \qquad (14)$$

Im allgemeinen ist $y_{t-1} \neq y_{-1,t-1}$, da die Mittelwerte $\overline{Y} \neq \overline{Y}_{-1}$.

Für das konstante Glied gilt:

$$\alpha = \overline{Y} - \lambda\, \overline{Y}_{-1} - \sum_{i=1}^{M} \beta_i\, \overline{X}_i \qquad (15)$$

Die Abweichungen vom Mittelwert können natürlich auch in Matrixschreibweise dargestellt werden. Die Zählung und Indizierung der Komponenten soll mit 1 in jedem Vektor begonnen werden, obwohl dann einander zeitlich entsprechende Beobachtungen nicht denselben Index erhalten. Die Rechnung wird jedoch wesentlich vereinfacht.

$$y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_{N-1} \end{bmatrix}, \quad y_{-1} = \begin{bmatrix} y_{-1,1} \\ \\ \\ y_{-1,N-1} \end{bmatrix} \quad x = \underset{[(N-1)xM]}{\begin{bmatrix} x_{11} & x_{M1} \\ & \\ & \\ x_{1,N-1} & x_{M,N-1} \end{bmatrix}} \quad \beta = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix} \tag{16}$$

Gleichung (5) lautet nun in Abweichungsform und Matrixschreibweise:

$$\begin{aligned} y &= \lambda y_{-1} + x\beta + v \\ v &= \varepsilon - \lambda \varepsilon_{-1} \end{aligned} \tag{17}$$

$$v = \begin{bmatrix} v_1 \\ \cdot \\ \cdot \\ v_{N-1} \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{N-1} \end{bmatrix} \qquad \varepsilon_{-1} = \begin{bmatrix} \varepsilon_{-1,1} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{-1,N-1} \end{bmatrix}$$

Zu minimieren ist die Summe (Lit.4)

$\varepsilon'\varepsilon + \varepsilon'_{-1}\ \varepsilon_{-1} \Longrightarrow$ Minimum bezüglich $\beta, \eta_{-1}, \lambda$

wobei $\eta_{-1} = y_{-1} - \varepsilon_{-1}$.

$$\begin{aligned} \varepsilon &= y - \lambda\eta_{-1} - x\beta \\ \varepsilon_{-1} &= y_{-1} - \eta_{-1} \\ \varepsilon'\varepsilon &= (y - \lambda\eta_{-1} - x\beta)'(y - \lambda\eta_{-1} - x\beta) \\ \varepsilon'_{-1}\varepsilon_{-1} &= (y'_{-1} - \eta'_{-1})(y_{-1} - \eta_{-1}) \end{aligned} \tag{18}$$

Die Abweichungen vom Mittelwert können natürlich auch in Matrixschreibweise dargestellt werden. Die Zählung und Indizierung der Komponenten soll mit 1 in jedem Vektor begonnen werden, obwohl dann einander zeitlich entsprechende Beobachtungen nicht denselben Index erhalten. Die Rechnung wird jedoch wesentlich vereinfacht.

$$y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_{N-1} \end{bmatrix}, \quad y_{-1} = \begin{bmatrix} y_{-1,1} \\ \\ \\ y_{-1,N-1} \end{bmatrix} \quad x = \underset{[(N-1)xM]}{\begin{bmatrix} x_{11} & x_{M1} \\ & \\ x_{1,N-1} & x_{M,N-1} \end{bmatrix}} \quad \beta = \begin{bmatrix} \beta_1 \\ \cdot \\ \cdot \\ \beta_M \end{bmatrix} \tag{16}$$

Gleichung (5) lautet nun in Abweichungsform und Matrixschreibweise:

$$\begin{aligned} y &= \lambda y_{-1} + x\beta + v \\ v &= \varepsilon - \lambda \varepsilon_{-1} \end{aligned} \tag{17}$$

$$v = \begin{bmatrix} v_1 \\ \cdot \\ \cdot \\ v_{N-1} \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{N-1} \end{bmatrix} \qquad \varepsilon_{-1} = \begin{bmatrix} \varepsilon_{-1,1} \\ \cdot \\ \cdot \\ \cdot \\ \varepsilon_{-1,N-1} \end{bmatrix}$$

Zu minimieren ist die Summe (Lit.4)

$$\varepsilon'\varepsilon + \varepsilon'_{-1}\,\varepsilon_{-1} \Longrightarrow \text{Minimum bezüglich} \quad \beta, \eta_{-1}, \lambda$$

wobei $\eta_{-1} = y_{-1} - \varepsilon_{-1}$.

$$\begin{aligned} \varepsilon &= y - \lambda\eta_{-1} - x\beta \\ \varepsilon_{-1} &= y_{-1} - \eta_{-1} \\ \varepsilon'\varepsilon &= (y - \lambda\eta_{-1} - x\beta)'(y - \lambda\eta_{-1} - x\beta) \\ \varepsilon'_{-1}\varepsilon_{-1} &= (y'_{-1} - \eta'_{-1})(y_{-1} - \eta_{-1}) \end{aligned} \tag{18}$$

Die Minimierung erfolgt durch Nullsetzen der Ableitungen nach den Komponenten von β und η_{-1} und durch Ableitung nach λ.

Die Ableitung nach $\eta_{-1,t}$ ergibt:

$$\eta_{-1} = \frac{1}{1+\lambda^2}(y_{-1} + \lambda y - \lambda x\beta) \tag{19}$$

Einsetzen von (19) in (18) führt zu

$$\varepsilon'\varepsilon + \varepsilon'_{-1}\,\varepsilon_{-1} = \frac{1}{1+\lambda^2}\left[y - \lambda y_{-1} - x\beta\right]'\left[y - \lambda y_{-1} - x\beta\right] \tag{20}$$

Partielle Ableitung nach β' ergibt

$$-x'y + \lambda x'y_{-1} + x'x\beta = 0 \tag{21}$$

oder, wenn $(x'x)$ regulär ist

$$\beta = (x'x)^{-1}(x'y - \lambda x'\,y_{-1}) \tag{22}$$

Partielle Ableitung von (20) nach λ führt zur zweiten Gleichung zur Bestimmung von β und λ :

$$\begin{aligned} &\lambda^2\left[y'y_{-1} - \beta'x'y_{-1}\right] + \\ &+ \lambda\left[2\,\beta'x'y - y'y - \beta'x'x\beta + y'_{-1}\,y_{-1}\right] + \\ &+ 1\left[-y'y_{-1} + \beta'x'y_{-1}\right] = 0 \end{aligned} \tag{23}$$

Wird (22) in (23) eingesetzt, erhält man eine quadratische Gleichung in λ (die Glieder dritten Grades verschwinden), mit folgendem Aussehen:

$$\lambda^2 \left[y' \, (I - x' \, (x'x)^{-1} x) y_{-1}\right] \; +$$

$$+ \lambda \left[y_{-1}' \, (I - x'(x'x)^{-1} x) y_{-1} - y' \, (I - x'(x'x)^{-1} x) \, y\right] \; +$$

$$+ 1 \left[- y' \, (I - x'(x'x)^{-1} x) \, y\right] = 0 \qquad (24)$$

oder wenn $\quad A = I - x'(x'x)^{-1} x$

$$\lambda^2 \, (y'Ay_{-1}) + \lambda (y_{-1}' \, Ay_{-1} - y'Ay) - y'Ay_{-1} = 0 \qquad (25)$$

Diese Gleichung besitzt 2 zueinander negativ reziproke Lösungen in λ .
Die ökonomische Praxis beschränkt uns im allgemeinen auf die Lösung $|\lambda_1| < 1$, sonst würde sich eine ökonomische Variable jetzt umso stärker auswirken, je weiter sie von der Gegenwart entfernt ist.
Der Schätzwert $\hat{\lambda}$ wird in (22) eingesetzt und ergibt $\hat{\beta}$.
$\hat{\beta}$ und $\hat{\lambda}$, in (19) eingesetzt, haben $\hat{\eta}_{-1}$ als Ergebnis.

Ausgewählte Literatur

1. Zvi GRILICHES: Distributed Lags (A Survey), Econometrica 1967, Vol. 35, p.16.
2. L.M.KOYCK: Distributed Lags and Investment Analysis, North Holland, Amsterdam, 1954.
3. R.N. WAUD: Misspecification in the "Partial Adjustment" and "Adaptive Expectations" Models, Int. Ec.Review, Vol.9, Nr.2, Juni 1968, p. 204.
4. L.R. KLEIN: The Estimation of Distributed Lags, Econometrica 1958, Vol. 26, p. 553.
5. T.J. SARGENT: Some Evidence on the Small Sample Properties of Distribution Lag Estimators in the Presence of Autocorrelated Disturbances, Review of Economics and Statistics 1968, Vo. 50, p. 87.
6. M. KENDALL, A. STUART: The Advanced Theory of Statistics, Vol. 2, Kap. 29.19, pp. 386-387, Charles Griffin & Co., London 1961.
7. T. AMEMIYA, W.A. FULLER: A Comparative Study of Alternative Estimators in a Distributed Lag Model, Econometrica 1967, Vol. 35, p. 517.
8. J. JOHNSTON: Econometric Methods, International Student Edition, McGraw-Hill, New York 1960, p. 217.